Deep Image Denoising with Adaptive Priors

Bo Jiang, Yao Lu*, Jiahuan Wang, Guangming Lu*, and David Zhang, *Life Fellow, IEEE*

***Abstract*—Image denoising methods using deep neural networks have achieved a great progress in the image restoration. However, the recovered images restored by these deep denoising methods usually suffer from severe over-smoothness, artifacts, and detail loss. To improve the quality of restored images, we first propose Supplemental Priors (SP) method to adaptively predict depth-directed and sample-directed prior information for the reconstruction (decoder) networks. Furthermore, the over-parameterized deep neural networks and too precise supplemental prior information may cause an over-fitting, restricting the performance promotion. To improve the generalization of denoising networks, we further propose Regularization Priors (RP) method to flexibly learn depth-directed and dataset-directed regularization noise for the retrieving (encoder) networks. By respectively integrating the encoder and decoder with these plug-and-play RP block and SP block, we propose the final Adaptive Prior Denoising Networks, called APD-Nets. APD-Nets is the first attempt to simultaneously regularize and supplement denoising networks from the adaptive priors' view with drawing learning-based mechanism into producing adaptive regularization noise and supplemental information. Extensive experiment results demonstrate our method significantly improves the generalization of denoising networks and the quality of restored images with greatly outperforming the traditional deep denoising methods both quantitatively and visually. *The code will be released at https://github.com/JiangBoCS/APD-Nets*.**

***Index Terms*—Image denoising, Over-smoothness, Over-fitting, Adaptive priors, Regularization priors, Supplemental priors.**

I. Introduction

Image denoising aims to remove noise from the noisy images to recover high-quality images. With the development of deep learning, most deep image denoising methods using Deep Neural Networks (DNNs) have significantly outperformed the non-deep denoising methods [1], [2], [3], [4]. Thus, this paper will focus on the deep denoising methods. However, the restored images usually suffer from grievous over-smoothness, artifacts, and detail loss. In the non-deep denoising methods, many methods based on designing supplemental prior information from experience and knowledge have been proposed. For example, the prior information can be achieved through non-linear total variation [5], bilateral variance estimation [6], weighted nuclear norm minimization [7], trilateral weighted sparse coding [8], and non-local self-similarity [9], [10] methods. However, these manual-design methods are not flexible and effective enough when the images' noise distributions are highly various and complex. Moreover, due to the inflexibility of these methods, they can not be directly integrated into the deep denoising networks. Accordingly, it is pressing to propose an adaptive prior method for the deep denoising methods.

To improve the quality of restored images, we propose Supplemental Priors (SP) method from the adaptive priors' view to predict depth-directed and sample-directed supplemental prior information for the reconstruction (decoder) stage in deep denoising networks. Using the features learned from deep neural networks, the prior information [11] of the specific task can be computed and supplemented to the networks to improve the performance. Inspired by this, we supplement the detail prior information learned from the sample feature and the networks' structure information for recovering images. Hence, such supplemented prior information is called Supplemental Priors (SP). Since the decoder networks are in the hierarchical structure, the layers in different depths have various reconstruction extents (reconstruction levels) in recovering images. Hence, the information of reconstruction extent can guide the SP method to generate depth-directed prior information. Owing to the difficult representation of such reconstruction extent, the proposed SP method first uses a learnable and low-cost embedded matrix, called External Reconstruction Extent (ERE), to represent it. The learned ERE is then used as the partial external information to generate the final depth-directed supplemental prior information. We jointly utilize the SSA and SCA modules to retrieve the statistical feature for every sample. Then, such statistical feature is used as the partial information in SP to produce the final sample-directed supplemental information. Finally, the supplemental prior information is produced through incorporating ERE, SCA and SSA in the SP block. Such supplemental prior information is much more appropriate and adaptive both on the depths and samples, producing high-quality restored images.

Furthermore, the DNNs are usually over-parameterized, easily causing an over-fitting in networks [12]. In addition, due to the powerful learning ability of the proposed SP method, the predicted supplemental prior information may be too precise for some samples, also probably leading to the over-fitting in networks. Such over-fitting will restrict the generalization of deep denoising methods, and thus hinder the performance promotion. Current proposed regularization methods to alleviate the over-fitting in DNNs can be classified into two categories. The first kind of methods are augmenting the training samples or training labels. For example, traditional augmentation [13] and Cutout [14] mainly augment the training samples, while Mixout [15], CutMix [16] and Maxup [17] mainly augment both the training samples and labels. However, these augmentation methods require multiple times of training cost, and the way of augmenting labels can not be reasonably applied to the denoised labels/images. The second kind of methods are augmenting in the feature space of DNNs. For example, Shake-shake [18] and Shake-Drop [19] methods inject

*Guangming Lu and Yao Lu are corresponding authors.

Bo Jiang, Yao Lu, Jiahuan Wang and Guangming Lu are with the Department of Computer Science and Technology, Harbin Institute of Technology at Shenzhen, Shenzhen 518057, China (e-mail: jiangbo_PhD@outlook.com; luyao2021@hit.edu.cn; wangjiahuanszhit@163.com; luguangm@hit.edu.cn).

David Zhang is with the School of Data Science, The Chinese University of Hong Kong (Shenzhen), China (e-mail: davidzhang@cuhk.edu.cn)

regularization noise into each layer in networks. However, these methods can only be utilized in Residual family networks [20], [21], because the residual connections can prevent learning perturbation during the regularization process. Furthermore, the manual-design noise in the above methods heavily depends on the prior experiences, leading to poor flexibility and generalization for different DNNs and datasets. From the above analyses, current regularization methods are not practical enough for deep denoising networks.

To improve the generalization of deep denoising methods, we further propose Regularization Priors (RP) method for the retrieving (encoder) stage in deep denoising networks. Since the encoder networks for retrieving features are also hierarchical, the layers in different depths will catch the features with various captured extents (captured levels). Hence, the captured extents can guide our RP method to generate pertinent noise for different depths. Similarly, the proposed RP method also employs a learnable embedded matrix to represent the External Captured Extent (ECE). Thus, the learned ECE can be utilized as the partial information to generate the depth-directed regularization noise. Similarly, the proposed RP method also employs a learnable embedded matrix to represent the External Captured Extent (ECE). Then, the learnable noise distribution is equipped in RP to generate batch noise for every input batch. Finally, RP method produces the regularization noise by integrating ECE with batch noise. Such regularization noise is finer and more adaptive towards the networks' depths. Furthermore, since the noise distributions are optimized according to the data distribution of entire dataset, the generated regularization noise is adaptive towards datasets and can be flexibly adjusted based on different datasets. Hence, the proposed RP method is general for various datasets and practical for deep denoising networks.

Respectively integrating the encoder and decoder with the proposed plug-and-play RP and SP blocks, we introduce the Adaptive Prior Denoising Networks, called APD-Nets (Fig.1). Jointly utilizing the adaptive regularization and supplemental priors, the proposed APD-Nets can effectively improve the generalization of networks and the quality of restored images. In summary, this paper mainly brings the following contributions:

- To improve the quality of restored images, we first propose Supplemental Prior (SP) method for the reconstruction (decoder) stage in deep denoising networks. SP integrates the learnable External Reconstruction Extent (ERE) with Semantic Channel Attention (SCA) and Semantic Spatial Attention (SSA) to predict depth-directed and sample-directed supplemental prior information. Such supplemental prior information is more adaptive and precise for recovering images.
- To improve the generalization of denoising networks, we further propose Regularization Priors (RP) method for the retrieving (encoder) stage in deep denoising networks. RP block integrates the learnable External Captured Extent (ECE) with learnable noise distributions to predict depth-directed and dataset-directed regularization noise. Such regularization noise is more general towards various datasets and more practical for deep denoising networks.
- Respectively inserting the RP and SP blocks into the encoder and decoder, Adaptive Prior Denoising Networks (APD-Nets) is proposed. *This is the first attempt to simultaneously regularize and supplement denoising networks from adaptive priors' view with drawing learning-based mechanism into producing adaptive regularization noise and supplemental information.* Extensive experiment results prove our method markedly improves the generalization of networks and the quality of restored images.

The paper is organized as follows: Section II introduces the related works. Sections III presents the proposed Adaptive Prior Denoising Networks. Section IV demonstrates the experiment results, and the paper is finally concluded in Section V.

II. Related Work

With the development of DNNs, deep denoising methods have achieved unprecedented performance in the image denoising area. Some works [22], [23], [24], [25], [26] respectively utilized various designed sub-structures in the networks to extract different information from the deep features. Such extracted information is beneficial to the denoising performance. DnCNN [27] introduced the residual learning to generate noise and achieved the end-to-end denoising. FFDNet [28] used the hyper-parameter of noise level to suppress non-uniform noise in the noisy images. RIDNet [29] used residual structure to reduce the flow of low-frequency information, and further applied the attention mechanism to enhance the interaction of channel features. However, these deep denoising methods still suffer from tough over-smoothness, artifacts, and detail loss in the restored images. Furthermore, the over-parameterized DNNs further cause an over-fitting. These two problems severely hinder the improvement of denoising performance.

To relieve the over-smoothness, artifacts, and detail loss, some prior information was designed and injected to the non-deep denoising methods. Such priors include external statistical prior [30], [31], sparsity [32], [8], and non-local self-similarity [33], [34], [35]. For example, NLM [33] and BM3D [34] used prior information of non-local self-similar patches to uncover self-similar patches in different conversion domains to remove noise. However, owing to the inflexibility of these manual-design priors, deep denoising networks can not directly employ them.

To relieve the over-fitting of networks, the first kind of regularization methods augmenting training samples or training labels have been developed. For example, traditional augmentation [36] cropped, flipped, and rotated images; Cutout [14] blurred the images with zero noise in the rectangular areas of images; Mixout [17] simultaneously mixed every pair of images and labels; Cutmix [16] and Maxup [17] are the extended versions of integrating Cutout and Mixout. These regularization methods are widely used in the high-level computer vision tasks, such as classification, detection, and segmentation. However, for the low-level image denoising methods, augmentation methods are not practical because the denoised labels/images can not be reasonably augmented.

Another kind of regularization methods are augmenting in the feature space of DNNs. Shake-Shake [18] generated

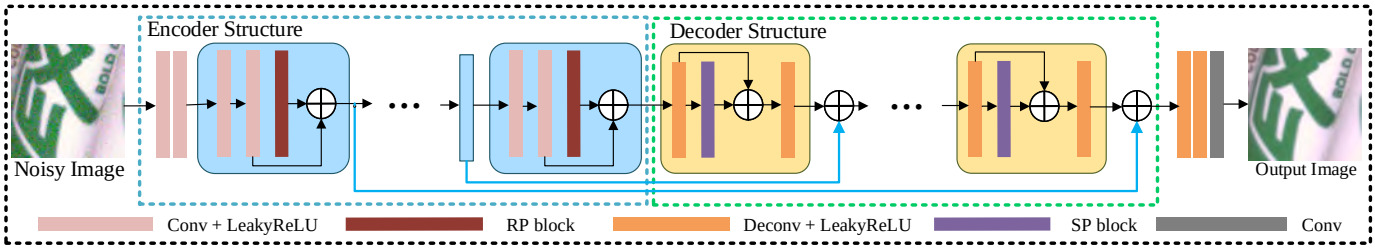


Fig. 1. The framework of APD-Nets. APD-Nets includes two main components, *i.e.*, encoder with RP block and decoder with SP block. In the retrieving feature process, the Regularization Prior (RP) block is inserted into every block of the encoder networks. In the reconstructing image process, the Supplemental Prior (SP) block is inserted into every block of the decoder networks. $\bigoplus$ indicates the element-wise addition.

regularization noise from manual-design noise distributions for the multi-branch networks; ShakeDrop [19], a generalized version of Shake-Shake, can be used in the single-branch networks. Although these regularization methods do not require the augmented labels, they should be used in the residual networks, because the residual connections can decrease the learning perturbation in the regularization process. Besides this, the regularization performance is sensitive to the manual-design noise distributions. Hence, these two regularization methods are also not practical enough for the deep denoising methods.

Different from the traditional supplemental prior information, the proposed Supplemental Prior (SP) method uses the learning technology and draws the learnable External Reconstruction Extent (ERE), Semantic Channel Attention (SCA), and Semantic Spatial Attention (SSA) to adaptively predict the supplemental prior information towards various depths and samples. Different from the traditional deep regularization methods, the proposed Regularization Prior (RP) method utilizes the learning technology, learnable External Captured Extent (ECE), and learnable noise distributions to flexibly generate the regularization noise for various depths and datasets. Different from the traditional deep denoising networks, we propose the adaptive priors, *i.e.*, SP and RP, from the adaptive priors' view to simultaneously regularize and supplement denoising networks.

III. Adaptive Prior Denoising Networks

This paper explicitly learns and induces the adaptive priors for the denoising networks to regularize and supplement the networks. Inspired by the encoder-decoder image mapping scheme of U-Net [37], the proposed APD-Nets is also implemented based on the encoder-decoder structure. In the retrieving feature process, the Regularization Prior (RP) block is inserted into every block of the encoder networks. In the reconstructing image process, the Supplemental Prior (SP) block is inserted into every block of the decoder networks. The framework of the proposed APD-Nets is shown in Fig. 1. Specifically, the base block of encoder includes the convolutional layer, Leaky ReLU layer, and RP block, while the base block of decoder consists of the deconvolutional layer, Leaky ReLU layer and SP block.

A. Supplemental Priors Method

To improve the quality of restored images, we propose the Supplemental Priors (SP) method (Fig. 2) to adaptively predict depth-directed and sample-directed supplemental prior information for reconstructing restored images. The proposed SP method first employs the External Reconstruction Extent (ERE) to explicitly learn the reconstruction extent information according to the statistical characteristics of the feature maps in different depth layers. According to the generative idea [38], such information can guide the SP block to generate much finer supplemental prior information with depth pertinence. Since ERE is difficult to be represented only relying on the current understandings of DNNs, we use a low-cost learnable embedded matrix $\boldsymbol{p}$ ($\boldsymbol{p} \in \mathbb{R}^{1\times H\times W}$) to represent it, where H and W indicate the height and width of the feature map size, respectively. ERE can be optimized along with the entire training process of APD-Nets.

The SP method further jointly uses Semantic Channel Attention (SCA) and Semantic Spatial Attention (SSA) to comprehensively exploit guidance information for every input sample. The structure of SCA is implemented based on the Squeeze-and-Excitation attention [39] to generate channel attention, illustrated in branch I in Fig. 2. In the SCA, a global average pooling layer first retrieves the global spatial information from the input features. Suppose $\boldsymbol{x} = [x_1, \cdots, x_c, \cdots, x_C]$ is the input tensor, where $x_c \in \mathbb{R}^{1\times H\times W}$. C, H and W denote the channels, height and width of the feature maps, respectively. Thus, the intermediate statistical channel information $\boldsymbol{k} = [k_1, \cdots, k_c, \cdots, k_C]$ is denoted by:

$$k_c = F_{GAP}(x_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j), \tag{1}$$

where $x_c(i,j)$ is the value at position (i,j) of x_c. $F_{GAP}(\cdot)$ denotes the global average pooling operation. Then, another two fully connection layers are attached. The final SCA attention A_{sca} is obtained:

$$A_{sca} = \delta^s(F_{f_C}(\delta^r(F_{f_{C/r}}(\boldsymbol{k})))), \tag{2}$$

where F_{f_C} and $F_{f_{C/r}}$ denote the fully connection layers with C and C/r ($r \in \{1, ..., C\}$) output channels, respectively. δ^s and δ^r indicate the activation Sigmoid layer and ReLU layer, respectively.

The proposed Semantic Spatial Attention (SSA), shown in branch II in Fig. 2, can generate pixel-wise attention at the spatial axis of feature maps. We first utilize the "CNR" module and max pooling layer to retrieve the correlations within the nearest pixels of the summation features from the encoder and

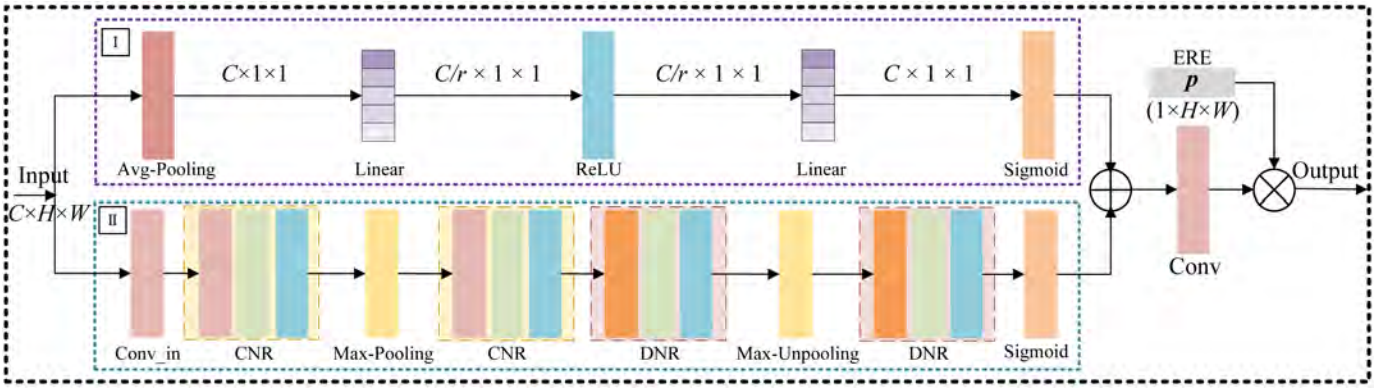


Fig. 2. The structure of Supplemental Prior (SP) block. The SP block includes External Reconstruction Extent (ERE), Semantic Channel Attention (SCA) branch (I), and Semantic Spatial Attention (SSA) branch (II). "CNR" stands for the convolutional, Normalization and ReLU layers in series. "DNR" represents the deconvolutional, Normalization and ReLU layers in series. $\oplus$ indicates the element-wise addition. $\otimes$ indicates the element-wise multiplication.

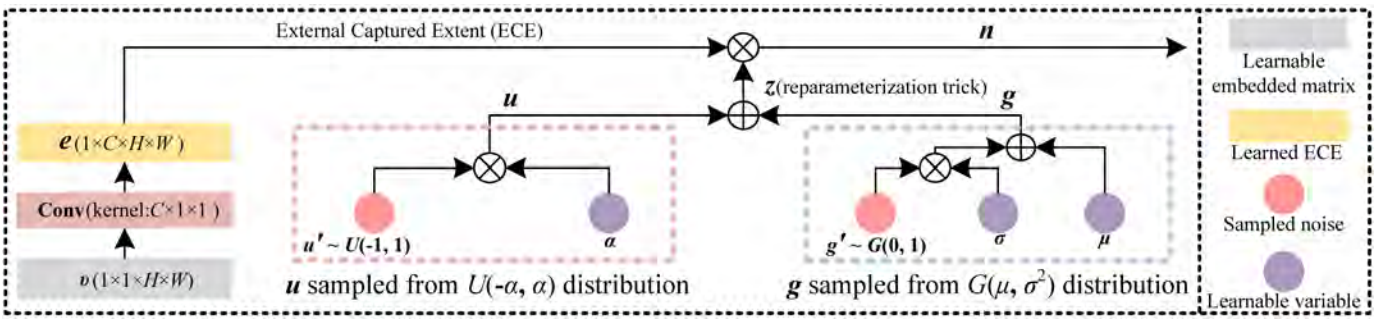


Fig. 3. The framework of Regularization priors (RP) block. RP block includes External Captured Extent (ECE) and learnable noise sampled from Gaussian and Uniform noise distributions. B, C, H and W indicate the batch size, channels, height and width, respectively. $\oplus$ indicates the element-wise addition. $\otimes$ indicates the element-wise multiplication.

decoder, where "CNR" is the combination of convolutional layer, normalization layer and ReLU layer. Then, we utilize "DNR" module and max un-pooling layer to generate pixel-wise attention for feature maps, where "DNR" is the combination of deconvolutional layer, normalization layer and ReLU layer. Thus, the final SSA attention A_{ssa} is formulated by:

$$A_{ssa} = \delta^s\left(F^2_{DNR}\left(F^1_{DNR}\left(F^2_{CNR}\left(F^1_{CNR}\left(F_{in}\left(\boldsymbol{x}\right)\right)\downarrow\right)\right)\uparrow\right)\right), \tag{3}$$

where F_{in} is the convolutional layer to transform the input feature. F^1_{CNR} and F^2_{CNR} are the "CNR" layers, while F^1_{DNR} and F^2_{DNR} are the "DNR" layers. "$\downarrow$" and "$\uparrow$" respectively indicate the max pooling layer and max un-pooling layer with scale factor of 2, respectively. Next, another convolutional layer is employed to fuse the SCA and SSA, generating comprehensive attention A. Finally, the proposed SP method can predict the final supplemental prior information $\boldsymbol{s}$ by multiplying ERE ($\boldsymbol{p}$) and comprehensive attention A. $\boldsymbol{s}$ is formulated as below:

$$\boldsymbol{s} = A \times \boldsymbol{p} = F_C(A_{sca} + A_{ssa}) \times \boldsymbol{p}, \tag{4}$$

where F_C denotes the last convolutional layer with C output channels, A_{sca} and A_{ssa} are added element by element based on the broadcast mechanism[1]. Through the above demonstrations, the proposed SP method can adaptively predict the pertinent and appropriate prior information according to various depths and samples. This is greatly beneficial for improving the quality of recovered images. The forward procedure of the proposed SP block is shown in Algorithm 1.

Algorithm 1: Forward procedure of the SP block.

Initialize:
  SP block's parameters : $\boldsymbol{\theta}_{SP}$,
  Learnable matrix ERE : $\boldsymbol{p} \in \mathbb{R}^{1\times C\times H\times W}$, $\boldsymbol{p} \sim \mathcal{N}(0,1)$.
Forward:
  Input: $\boldsymbol{x} \in \mathbb{R}^{1\times C\times H\times W}$, and $\boldsymbol{p}$,
  Generate SCA attention A_{sca} according to Eqn. 2,
  Generate SSA attention A_{ssa} according to Eqn. 3,
  Compute supplemental prior $\boldsymbol{s}$ according to Eqn. 4,
  Output: Supplemental prior $\boldsymbol{s}$.

B. Regularization Priors Method

To improve the generalization of deep denoising networks, the Regularization Priors (RP) method is proposed. Inspired by the idea of regularization based on additive noise [5], the proposed RP method can generate depth-directed and dataset-directed regularization noise through integrating the learnable

[1]https://numpy.org

Algorithm 2: Forward procedure of the RP block.

Initialize:
  RP block's parameters : $\boldsymbol{\theta}_{RP}$,
  Parameters of noise distributions:
  $\boldsymbol{g} \sim G(\mu, \sigma^2), \boldsymbol{u} \sim U(-\alpha, \alpha),\ \boldsymbol{\rho} = \{\mu, \sigma, \alpha\}$,
  Learnable matrix in ECE : $\boldsymbol{v} \in \mathbb{R}^{1\times1\times H\times W}$,
  $\boldsymbol{v} \sim \mathcal{N}(0, 1)$.
Forward:
  Input: Sampled noise $\boldsymbol{g'} \sim G(0, 1)$ and
  $\boldsymbol{u'} \sim U(-1, 1)$ and $\boldsymbol{v}$,
  Generate ECE matrix $\boldsymbol{e}$ according to Eqn. 5,
  Produce mixed sampled noise $\boldsymbol{z}$ according to Eqn. 7,
  Compute regularization noise $\boldsymbol{n}$ according to Eqn. 8,
  Output: Regularization noise $\boldsymbol{n}$.

External Captured Extent (ECE, $\boldsymbol{e}$) and learnable mixed noise $\boldsymbol{z}$ (tensor). The framework of the RP block is illustrated in Fig. 3. Specifically, the RP block generates ECE ($\boldsymbol{e} \in \mathbb{R}^{1\times C\times H\times W}$) through expanding and transforming a learnable embedded matrix $\boldsymbol{v}$ ($\boldsymbol{v} \in \mathbb{R}^{1\times1\times H\times W}$) utilizing a "$1 \times 1$" convolutional layer $F_{C_{1\times1}}(\cdot; \boldsymbol{\theta}_{RP})$, where $\boldsymbol{\theta}_{RP}$ denotes the parameters. Thus, ECE is formulated as below:

$$\boldsymbol{e} = F_{C_{1\times1}}(\boldsymbol{v}; \boldsymbol{\theta}_{RP}). \tag{5}$$

The learnable mixed noise is generated from the learnable mixed noise distributions. Traditionally, the noise is generally sampled from Gaussian noise distribution $G(\mu, \sigma^2)$, where μ and σ indicate the mean and standard deviation, respectively. In order to improve the learning disturbance of regularization process, the noise sampled from Uniform distribution $U(-\alpha, \alpha)$ ($\alpha > 0$) is supplemented to Gaussian noise. Thus, the final mixed sampled noise $\boldsymbol{z}$ is formulated as below:

$$\begin{gathered}\boldsymbol{z} = \boldsymbol{g} + \boldsymbol{u};\\ \text{where } \boldsymbol{g} \sim G(\mu, \sigma^2), \boldsymbol{u} \sim U(-\alpha, \alpha).\end{gathered} \tag{6}$$

In the training process, the sampling operation from G and U is not differentiable, leading to no gradients for the parameters μ, σ and α in the back-propagation stage. Thus, the reparameterization trick [40] in VAEs is employed to resolve this problem. The sampling operation is replaced with sampling from $G(0, 1)$ and $U(-1, 1)$. Then, such sampled noise is transformed using μ, σ and α. Hence, Eqn. 6 is reformulated as the following equation:

$$\begin{gathered}\boldsymbol{z} = (\boldsymbol{g'} \times \sigma + \mu) + (\boldsymbol{u'} \times \alpha);\\ \text{where } \boldsymbol{g'} \sim G(0, 1), \boldsymbol{u'} \sim U(-1, 1).\end{gathered} \tag{7}$$

The size of $\boldsymbol{g'}$ and $\boldsymbol{u'}$ in Eqn. 7 is $B \times 1 \times 1 \times 1$, where B indicates the number of training samples in every input batch. Next, the final regularization noise $\boldsymbol{n}$ is generated through multiplying the ECE and sampled mixed noise:

$$\begin{gathered}\boldsymbol{n}(\boldsymbol{v}, \boldsymbol{\theta}_{RP}, \boldsymbol{\rho}) = \boldsymbol{e}(\boldsymbol{v}, \boldsymbol{\theta}_{RP}) \times \boldsymbol{z}(\mu, \sigma, \alpha);\\ \text{where } \boldsymbol{\rho} = \{\mu, \sigma, \alpha\}.\end{gathered} \tag{8}$$

Through the above process, the proposed RP method can flexibly produce the regularization noise for different depths. Furthermore, although the regularization noise is generated without applying input sample's information, the parameters of RP block are optimized based on the entire data distribution of training set. Therefore, our RP block can also adjust the produced noise distribution on various datasets. This indicates the proposed regularization method is general for datasets with different data distributions. The forward procedure of the proposed RP block is shown in Algorithm 2.

Algorithm 3: Learning procedure of the APD-Nets.

Initialize:
  APD-Nets' parameters : $\boldsymbol{\theta}$,
  SP block's parameters : $\boldsymbol{\theta}_{SP}$,
  RP block's parameters : $\boldsymbol{\theta}_{RP}$,
  Parameters of noise distributions:
  $\boldsymbol{g} \sim G(\mu, \sigma^2), \boldsymbol{u} \sim U(-\alpha, \alpha),\ \boldsymbol{\rho} = \{\mu, \sigma, \alpha\}$,
  Learnable matrix ERE : $\boldsymbol{p} \in \mathbb{R}^{1\times C\times H\times W}$,
  $\boldsymbol{p} \sim \mathcal{N}(0, 1)$.
  Learnable matrix in ECE : $\boldsymbol{v} \in \mathbb{R}^{1\times1\times H\times W}$,
  $\boldsymbol{v} \sim \mathcal{N}(0, 1)$.
Repeat:
  Input: Noisy image $\boldsymbol{x}_n$, $\boldsymbol{p}$, $\boldsymbol{v}$, and sampled noise,
  Extract shallow features: $I_{SF} = F_{conv}(\boldsymbol{x}_n)$,
  Generate regularized features through the encoder integrated with RP blocks according to Eqns. 5 - 8,
  Produce and reconstruct the denoised image $\boldsymbol{x}_d$ through the decoder integrated with SP blocks according to Eqns. 1 - 4,
  Compute loss of $\boldsymbol{x}_c$ and $\boldsymbol{x}_d$ according to Eqn. 10,
  Update: $\boldsymbol{\theta}$, $\boldsymbol{\theta}_{SP}$, $\boldsymbol{\theta}_{RP}$, $\boldsymbol{\rho}$, $\boldsymbol{p}$ and $\boldsymbol{v}$.
Until convergence

C. APD-Nets

Respectively inserting the SP and RP blocks into the encoder and the decoder networks, the proposed APD-Nets for image denoising is constructed. According to the previous illustrations and the minimum absolute error loss (*i.e.*, $L_1 = \frac{1}{N}\sum \|I_{in} - I_{traget}\|$ [47]), where N denotes the number of training samples, the final loss function of the proposed APD-Nets is formulated as below:

$$\begin{aligned}&L_1\left(\boldsymbol{x}_n, \boldsymbol{x}_c | \boldsymbol{\theta}, \boldsymbol{\theta}_{SP}, \boldsymbol{\theta}_{RP}, \boldsymbol{p}, \boldsymbol{\rho}, \boldsymbol{v}\right) = \\ &\qquad \frac{1}{NS}\sum_{i=1}^{N}\sum_{j=1}^{S}\left\|x_c^i - f\left(x_n^i | \boldsymbol{\theta}, \boldsymbol{\theta}_{SP}, \boldsymbol{\theta}_{RP}, \boldsymbol{p}, \boldsymbol{\rho}, \boldsymbol{v}\right)\right\|,\end{aligned} \tag{9}$$

where $\boldsymbol{\theta}$ denotes the parameters of backbone encoder and decoder in APD-Nets. $\boldsymbol{\theta}_{SP}$ and $\boldsymbol{\theta}_{RP}$ indicate SP block's parameters and RP block's parameters, respectively. The noise sampling times in the RP block (Eqn. 7) is S. From Eqn. 9, it is apparent that the training cost is much larger than the regular training because of the additional S sampling operations of the regularization noise. Since the number of training iterations is very large for DNNs, the parameters in the SP and RP blocks can be optimized well by sampling the regularization noise only once for every input sample. This contributes to

TABLE I

AVERAGE PSNR OF THE DENOISED GRAYSCALE IMAGES FROM SET12 AND BSD68 DATASETS. THE VALUES OF PSNR AND SSIM ARE POSITIVELY CORRELATED WITH VISUAL QUALITY, WHILE THE VALUE OF LPIPS IS NEGATIVELY CORRELATED WITH VISUAL QUALITY. C INDICATES THAT THE NUMBER OF CHANNELS FOR EACH CONVOLUTIONAL LAYER OR DECONVOLUTION LAYER IS FIXED AT 144 OR 64. THE COMPUTATIONAL EFFICIENCY (FLOPS) IS OBTAINED WHEN THE PATCH SIZE OF 256×256.

Methods	Parameters	Parameters	σ_n = 15						σ_n = 25						σ_n = 50					
			Set12			BSD68			Set12			BSD68			Set12			BSD68		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
BM3D [34]	-	-	32.39	0.786	0.0377	31.07	0.769	0.0515	29.95	0.755	0.0659	28.56	0.737	0.0872	26.71	0.694	0.1218	25.63	0.687	0.1456
MLP [41]	-	-	32.42	0.787	0.0374	31.31	0.773	0.0488	30.12	0.757	0.0636	28.74	0.739	0.0843	26.88	0.711	0.1183	25.84	0.696	0.1408
UDNet [42]	-	-	32.72	0.789	0.0347	31.57	0.774	0.0459	30.32	0.759	0.0609	29.13	0.744	0.0780	27.13	0.721	0.1133	26.21	0.701	0.1325
DnCNN [27]	35.36 G	0.56M	32.66	0.785	0.0353	31.62	0.778	0.0454	30.35	0.760	0.0605	29.22	0.743	0.0766	27.17	0.721	0.1125	26.29	0.703	0.1307
MemNet [43]	190.77 G	2.91 M	32.72	0.787	0.0347	31.57	0.772	0.0459	30.39	0.761	0.0600	29.21	0.741	0.0768	27.26	0.722	0.1108	26.29	0.702	0.1307
FFDNet [28]	12.72 G	0.85 M	32.77	0.789	0.0343	31.62	0.777	0.0454	30.44	0.763	0.0593	29.23	0.745	0.0765	27.35	0.723	0.1090	26.32	0.707	0.1301
VDN [44]	100.97 G	7.82 M	32.78	0.791	0.0342	31.64	0.773	0.0452	30.50	0.763	0.0586	29.28	0.746	0.0757	27.48	0.724	0.1065	26.46	0.709	0.1271
RIDNet [29]	96.96 G	1.50 M	32.91	0.793	0.0331	31.81	0.780	0.0434	30.43	0.762	0.0595	29.34	0.747	0.0748	27.32	0.723	0.1096	26.41	0.708	0.1281
SADNet [45]	130.13 G	4.77 M	32.78	0.792	0.0342	31.68	0.776	0.0441	30.45	0.763	0.0592	29.31	0.745	0.0752	27.67	0.726	0.1021	26.51	0.712	0.1293
AINDNet [46]	102.42 G	1.69 M	32.92	0.793	0.0331	31.69	0.775	0.0446	30.60	0.764	0.0573	29.26	0.745	0.0760	27.43	0.724	0.1075	26.32	0.707	0.1301
APD-Nets-C64 (Ours)	22.05 G	1.55 M	32.96	0.794	0.0327	31.79	0.779	0.0436	30.74	0.765	0.0555	29.35	0.748	0.0747	27.78	0.727	0.1009	26.58	0.711	0.1245
APD-Nets-C144 (Ours)	116.1 G	7.71 M	**33.46**	**0.801**	**0.0289**	**32.04**	**0.785**	**0.0410**	**31.20**	**0.771**	**0.0500**	**29.75**	**0.752**	**0.0688**	**28.41**	**0.735**	**0.0898**	**27.16**	**0.719**	**0.1127**

TABLE II

AVERAGE PSNR OF THE DENOISED COLOR IMAGES FROM SET12 AND CBSD68 DATASETS. THE VALUES OF PSNR AND SSIM ARE POSITIVELY CORRELATED WITH VISUAL QUALITY, WHILE THE VALUE OF LPIPS IS NEGATIVELY CORRELATED WITH VISUAL QUALITY. C INDICATES THAT THE NUMBER OF CHANNELS FOR EACH CONVOLUTIONAL LAYER OR DECONVOLUTION LAYER IS FIXED TO 144 OR 64. THE COMPUTATIONAL EFFICIENCY (FLOPS) IS OBTAINED WHEN THE PATCH SIZE OF 256×256.

Methods	FLOPs	Parameters	σ_n = 15						σ_n = 25						σ_n = 50					
			Set12			CBSD68			Set12			CBSD68			Set12			CBSD68		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
BM3D [34]	-	-	33.42	0.782	0.0349	32.23	0.779	0.0492	31.66	0.763	0.0638	30.28	0.745	0.0853	28.43	0.725	0.1199	27.36	0.712	0.1433
MLP [41]	-	-	33.55	0.788	0.0346	32.31	0.782	0.0465	31.83	0.774	0.0615	30.46	0.749	0.0824	28.60	0.729	0.1164	27.57	0.717	0.1385
UDNet [42]	-	-	33.64	0.788	0.0319	32.51	0.782	0.0436	32.03	0.782	0.0588	30.84	0.752	0.0761	28.85	0.729	0.1114	27.94	0.711	0.1302
DnCNN [27]	36.66 G	0.56M	33.69	0.791	0.0325	32.58	0.782	0.0431	32.06	0.781	0.0584	30.92	0.757	0.0747	28.89	0.736	0.1106	28.97	0.728	0.1284
MemNet [43]	192.07 G	2.91 M	33.76	0.792	0.0319	32.59	0.788	0.0436	32.10	0.777	0.0579	30.93	0.759	0.0749	28.98	0.740	0.1089	28.01	0.720	0.1284
FFDNet [43]	14.02 G	0.85 M	33.82	0.794	0.0315	32.66	0.790	0.0431	32.14	0.769	0.0572	30.94	0.762	0.0746	29.07	0.735	0.1071	28.04	0.721	0.1278
VDN [44]	101.27 G	7.82 M	33.87	0.800	0.0314	32.70	0.791	0.0429	32.15	0.775	0.0565	30.97	0.763	0.0738	29.20	0.731	0.1046	28.19	0.723	0.1248
RIDNet [29]	98.26 G	1.50 M	33.88	0.802	0.0303	32.73	0.794	0.0411	32.16	0.772	0.0574	30.99	0.765	0.0729	29.04	0.731	0.1077	28.14	0.722	0.1258
SADNet [45]	131.43 G	4.77 M	33.97	0.802	0.0314	32.75	0.796	0.0418	32.21	0.764	0.0571	31.02	0.767	0.0733	29.39	0.730	0.1002	28.23	0.716	0.1270
AINDNet [46]	103.72 G	1.69 M	34.00	0.803	0.0303	32.78	0.799	0.0423	32.31	0.770	0.0552	31.05	0.768	0.0741	29.15	0.745	0.1056	28.05	0.729	0.1278
APD-Nets-C64 (Ours)	23.25 G	1.55 M	34.26	0.816	0.0299	33.19	0.801	0.0413	32.45	0.787	0.0534	31.06	0.769	0.0728	29.50	0.749	0.0990	28.30	0.734	0.1222
APD-Nets-C144 (Ours)	117.2 G	7.71 M	**34.55**	**0.822**	**0.0261**	**33.54**	**0.804**	**0.0387**	**32.91**	**0.793**	**0.0479**	**31.46**	**0.774**	**0.0669**	**30.13**	**0.757**	**0.0879**	**28.88**	**0.741**	**0.1104**

no additional overhead of training. Thus, Eqn. 10 is further reformulated as below:

$$L_1\left(\boldsymbol{x}_n, \boldsymbol{x}_c|\boldsymbol{\theta}, \boldsymbol{\theta}_{SP}, \boldsymbol{\theta}_{RP}, \boldsymbol{p}, \boldsymbol{\rho}, \boldsymbol{\upsilon}\right) = \frac{1}{N}\sum_{i=1}^{N}\|x_c^i - f\left(x_n^i|\boldsymbol{\theta}, \boldsymbol{\theta}_{SP}, \boldsymbol{\theta}_{RP}, \boldsymbol{p}, \boldsymbol{\rho}, \boldsymbol{\upsilon}\right)\|. \tag{10}$$

Finally, the algorithm of the proposed APD-Nets is shown in Algorithm 3.

IV. EXPERIMENTS AND ANALYSES

A. Implementation and Initialization

The proposed APD-Nets has four down-samplings and four up-samplings with the same scale factor of 2 for the encoder and decoder, respectively. Further, the encoder and decoder contain four and three base blocks, respectively. The number of channels (C) for each convolutional layer or deconvolution layer is fixed to 64 or 144. The kernel size of the down-sampling convolutional layers and up-sampling deconvolutional layers are set to "6×6", while "3×3" for all the other convolutional layers. In terms of PSNR and SSIM [48], many image restoration tasks with L1 loss function achieve better performance than L2 loss [49]. Thus, we adopt L1 loss function in this work. Our method is evaluated on both the synthetic datasets and real noisy datasets. We employ BSD500 [50] and DIV2K [51] datasets with adding different levels of noise ($\sigma_n = 15, 25, 50$) to synthetic noisy images. For real noisy images, we use the cropped patches with size "256×256" sampled from the SIDD [52] dataset. We adopt random rotations of $90°, 180°, 270°$, and flip the images horizontally for data augmentation. Every input training batch of the proposed APD-Nets is 128 patches with size "256×256" from the training image pairs. We use Adam [53] optimizer, in which β_1 and β_2 is set to 0.9 and 0.999, respectively. The learning rate is 1×10^{-4}. Finally, the evaluation metrics are the PSNR, SSIM [48] and LPIPS [54].

B. Experiment Comparisons

The proposed method is compared with many state-of-the-art algorithms. To be fair in comparison, we produce their results using public-available implementations provided by the reported settings in the corresponding literatures. In the synthetic noisy images experiments, we evaluate the compared methods on the Set12 and BSD68 [55] datasets with adding different levels of noise ($\sigma_n = 15, 25, 50$). In the comparisons of real-world noisy images, the validation sets are SSID [52], DND [56], and Nam [57].

Synthetic grayscale noisy images. In the comparisons of grayscale noisy images, we compare the proposed APD-Nets with BM3D [34], MLP [41], UDNet [42], DnCNN [27], MemNet [43], FFDNet [28], VDN [44], RIDNet [29], SADNet [45], and AINDNet [46]. Table I shows the average results of PSNR, SSIM and LPIPS on Set12 and BSD68 datasets with three different noise levels. From this table, our APD-Nets-C64 achieves competitive results compared to the state-of-the-art AINDNet, while the FLOPs of the APD-Nets-C64

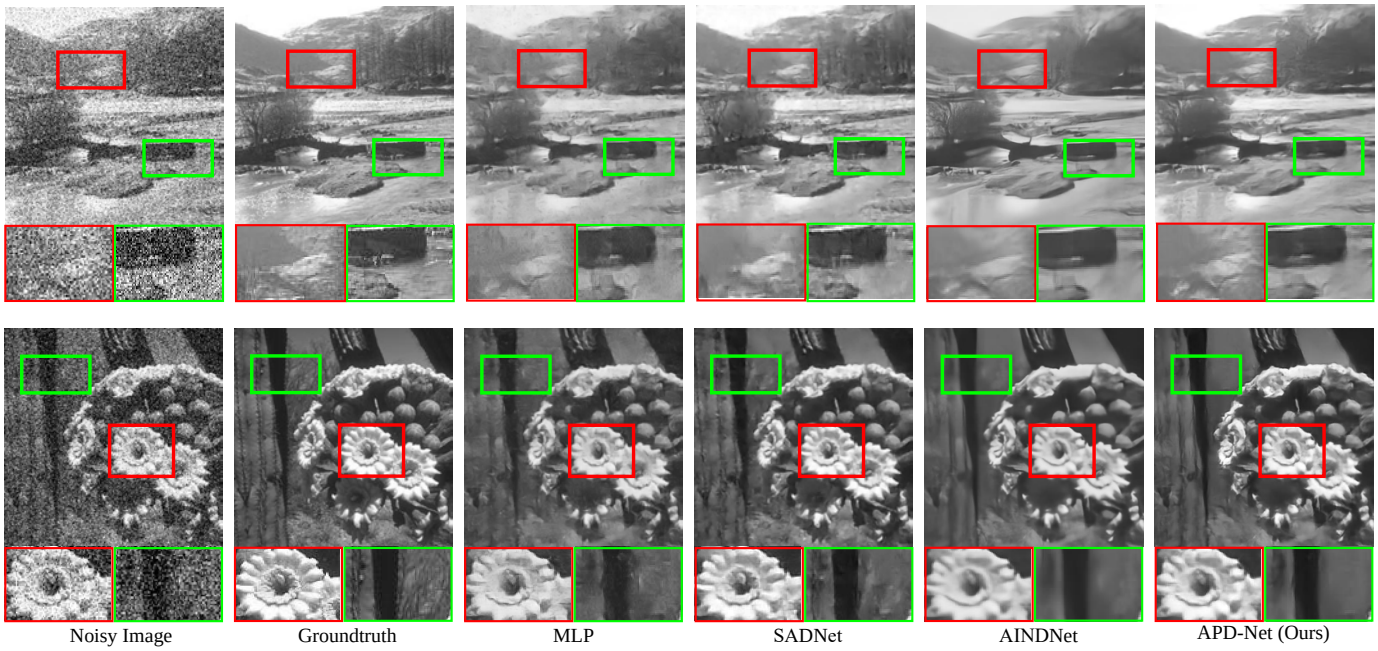


Fig. 4. Visual comparisons between APD-Net and its competitors (*i.e.*, MLP [41], SADNet [45] and AINDNet [46]) in the evaluation of grayscale-noisy image denoising. Both test images were from BSD68 with a large noise intensity of σ = 50. Please zoom in for a better view.

are reduced about **4.6 times** than AINDNets. In order to validate the capacity of our method, we construct the APD-Nets-C144 with approximately the same FLOPs of AINDNet. The performances produced by APD-Nets-144 are improved satisfactorily compared to AINDNet and are the best among all the compared methods. Additionally, more visual comparisons on BSD68 dataset are shown in Fig. 4. Compared with other methods, our method can recover images more satisfactorily without introducing apparent artifacts and over-smoothness. This indicates the effectiveness of the proposed SP and RP methods. Furthermore, since these test datasets are independent from the training datasets, these results fully show that the SP and RP methods in APD-Nets achieve a strong generalization for denoising the synthetic noisy grayscale images.

Synthetic color noisy images. Table II reports the quantitative results of all the compared methods. The proposed APD-Nets-C144 achieves the best performances on both Set12 and CBSD68 datasets. Particularly, compared to AINDNet [46], APD-Nets-C144 improves the average PSNR by 0.55 dB, 0.60 dB and 0.98 dB on Set12 dataset with $\sigma_n = 15$, $\sigma_n = 25$ and $\sigma_n = 50$, respectively. Also, APD-Nets-C144 promotes the average PSNR by 0.76 dB, 0.41 dB and 0.83 dB on CBSD68 dataset with $\sigma_n = 15$, $\sigma_n = 25$ and $\sigma_n = 50$, respectively. These comparisons demonstrate the effectiveness of the APD-Nets. Moreover, our APD-Nets-C64 produces better results than the VDN [44] on both Set12 and BSD68 datasets with $\sigma_n = 50$. However, the model size of the proposed APD-Nets-C64 is about **5 times** less than that of VDN. Moreover, our APD-Nets-C64 achieves better results compared to the state-of-the-art AINDNet, while the FLOPs of the APD-Nets-C64 are reduced about **4.4 times** than AINDNets. In order to validate the capacity of our method under the condition of the same computational complexity, we construct the APD-Nets-C144 with approximately the same FLOPs of AINDNet. The performances produced by APD-Nets-144 are improved satisfactorily compared to AINDNet and are the best among all the compared methods. This indicates our RP and SP methods can improve the learning ability of the denoising networks, leading to reducing the parameters to some extent. Fig. 5 shows the visual comparisons of different methods on CBSD68 dataset with $\sigma_n = 50$. The compared methods remove the noise while destroying the image details, resulting in over-smoothness, artifacts and severely smearing many texture areas. However, the visualization result achieved by the proposed APD-Nets is the closest to that of ground truth without introducing other artifacts. This is mainly because the proposed APD-Nets can adaptively generate precise supplemental prior information towards every depth and every input sample, leading to higher quality of the restored images.

Real-World Noisy Images. The real noise contained in real-world images is usually accumulated from multiple complex noise sources. Thus, denoising the real-world images is very practical but difficult. This implies it is persuasive to evaluate the denoising methods on the real-world noisy datasets. Table III reports the PSNR/SSIM/LPIPS results of different methods on DND, SSID, and Nam datasets. It is worth noting that, except for the MPRNet [59] and MIRNet [60] with much larger model size, our APD-Nets-C144 produce much better results in most evaluations. Also, the proposed APD-Nets-C144 generates competitive results compared to the MPRNet and MIRNet while using much fewer parameters. On the Nam dataset, the proposed APD-Nets-C144 improves the PSNR result by 0.24 dB compared to MIRNet. Furthermore, compared to the RIDNet [29] and AINDNet [46] with approximately the same model size, our APD-Nets-C64 improves the PSNR results by 0.13 dB and 0.37 dB on the SIDD dataset, respectively.

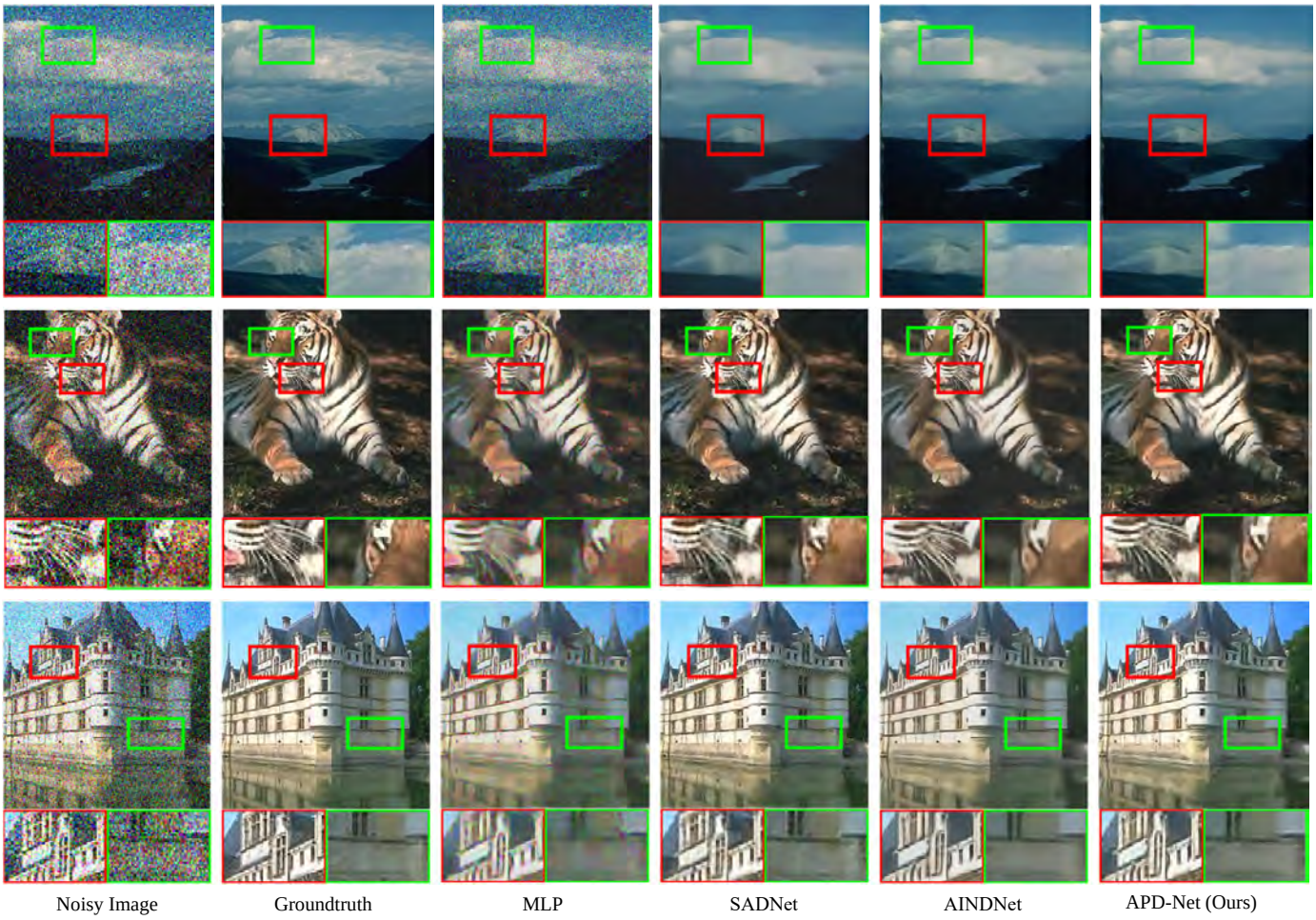


Fig. 5. Visual comparisons between APD-Net and its competitors (*i.e.*, MLP [41], SADNet [45] and AINDNet [46]) in the evaluation of color-noisy image denoising. All test images were from CBSD68 with a large noise intensity of $\sigma = 50$. Please zoom in for a better view.

TABLE III
AVERAGE PSNR/SSIM OF THE DENOISED REAL IMAGES FROM SIDD, DND AND NAM DATASETS. THE VALUES OF PSNR AND SSIM ARE POSITIVELY CORRELATED WITH VISUAL QUALITY, WHILE THE VALUE OF LPIPS IS NEGATIVELY CORRELATED WITH VISUAL QUALITY. THE COMPUTATIONAL EFFICIENCY (FLOPS) IS OBTAINED ON THE PATCH SIZE OF 256×256. THE INFERENCE SPEED IS OBTAINED WHEN THE PATCH SIZE OF 512×512.

Methods	Speed (s)	FLOPs	Parameters	SIDD			DND			Nam		
				PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
MLP [41]	0.089	–	–	24.71	0.641	0.1682	34.23	0.833	0.0239	34.63	0.852	0.0217
DnCNN [27]	0.064	36.66 G	0.56 M	23.66	0.583	0.1971	37.90	0.943	0.0134	34.95	0.885	0.0201
FFDNet [28]	0.028	14.02 G	0.85 M	34.19	0.857	0.0241	34.40	0.847	0.0229	38.81	0.957	0.0142
BDNet [58]	0.429	40.38 G	4.37 M	33.28	0.868	0.0302	38.06	0.942	0.0144	39.08	0.969	0.0137
RIDNet [29]	0.214	98.26 G	1.50 M	38.95	0.952	0.0147	39.23	0.953	0.0139	39.20	0.973	0.0139
AINDNet [46]	0.415	103.72 G	1.69 M	38.71	0.951	0.0145	39.37	0.951	0.0138	39.09	0.962	0.0142
SADNet [45]	0.358	131.43 G	4.77 M	39.46	0.957	0.0138	39.59	0.952	0.0136	39.89	0.979	0.0137
VDN [44]	0.497	101.27 G	7.82 M	39.23	0.971	0.0139	39.38	0.952	0.0137	39.68	0.976	0.0142
MPRNet [59]	1.058	588.14 G	15.74 M	39.71	0.958	0.0124	39.80	0.954	0.0133	39.97	0.981	0.0129
MIRNet [60]	2.082	555.54 G	31.78 M	**39.72**	**0.959**	0.0123	**39.88**	**0.956**	0.0120	40.03	0.986	0.0121
APD-Nets-C64 (Ours)	**0.015**	23.25 G	1.55 M	39.08	0.939	0.0137	39.21	0.948	0.0137	39.83	0.974	0.0132
APD-Nets-C144 (Ours)	0.167	117.2 G	7.71M	39.56	0.955	0.0129	39.69	0.952	0.0126	**40.27**	**0.987**	0.0118
APD-Nets-C224 (Ours)	0.853	283.25 G	18.61M	**39.75**	**0.963**	**0.0121**	**39.92**	**0.959**	**0.0112**	**40.36**	**0.989**	**0.0109**

Additionally, the inference time of our APD-Nets-C64 using the patch size of 512×512 is the smallest among all the evaluation methods. Furthermore, our APD-Nets-C144 achieves competitive results compared to the state-of-the-art MIRNet, while the FLOPs of the APD-Nets-C144 are reduced about **4.7 times** than MIRNet. To further verify the capacity of our method, we construct the APD-Nets-C224 with more parameters and FLOPs. The proposed APD-Nets-C224 achieves best performances. Furthermore, compared to the MIRNet, our APD-Nets-C224 significantly reduces about **2 times** parameters

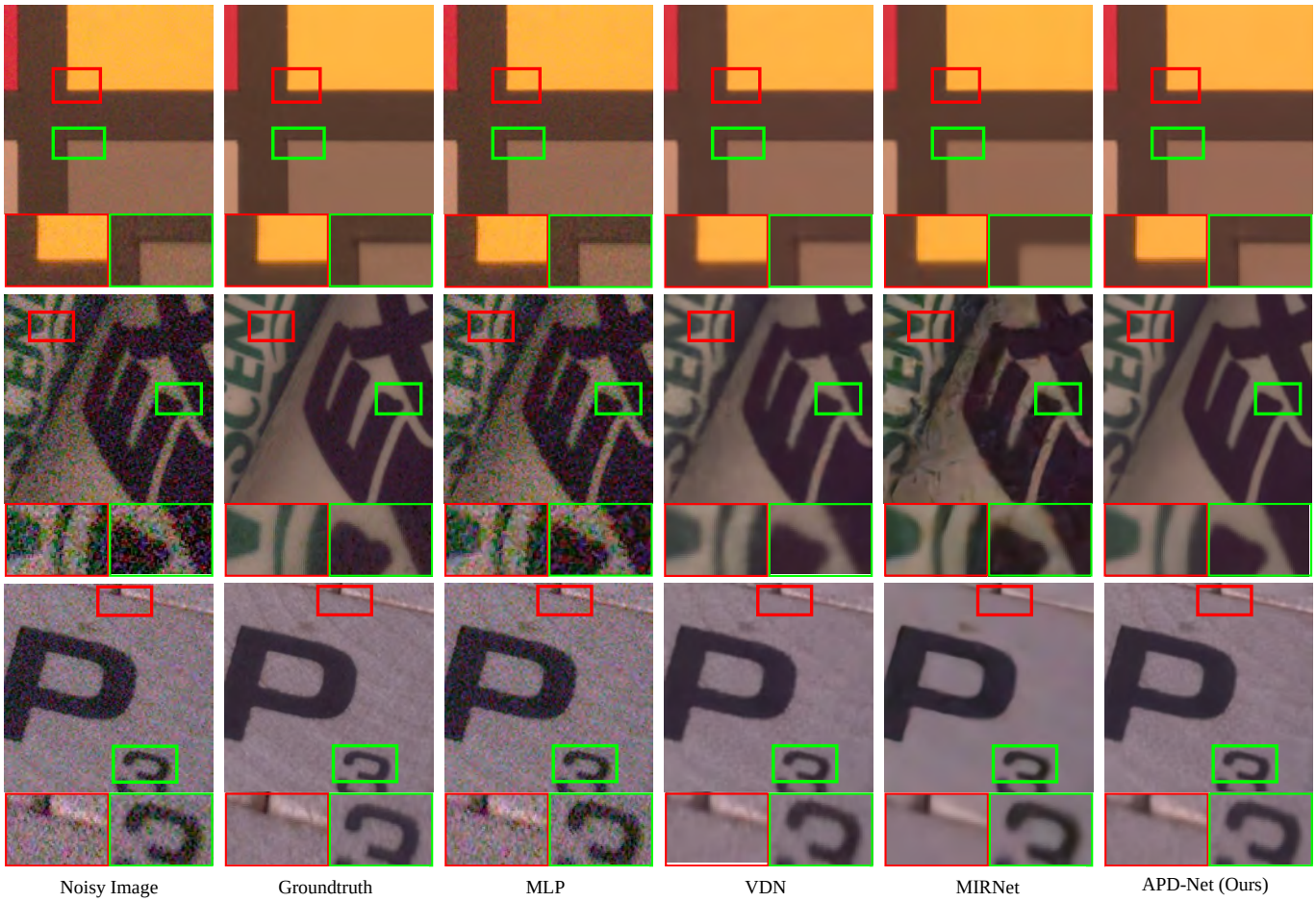


Fig. 6. Visual comparisons between APD-Net and its competitors (*i.e.*, MLP [41], VDN [44] and MIRNet [60]) in the evaluation of real-noisy image denoising. All test images were from real noisy dataset SIDD. Please zoom in for a better view.

TABLE IV
ABLATION STUDY OF DIFFERENT COMPONENTS. PSNR/SSIM VALUES ARE FROM SIDD DATASET. THE VALUES OF PSNR AND SSIM ARE POSITIVELY CORRELATED WITH VISUAL QUALITY, WHILE THE VALUE OF LPIPS IS NEGATIVELY CORRELATED WITH VISUAL QUALITY.

	Case	a	b	c	d	e	f	g	h	i	j
SP	ERE			✓	✓	✓	✓	✓	✓		✓
	SCA			✓	✓	✓	✓	✓		✓	✓
	SSA			✓	✓	✓	✓		✓	✓	✓
RP	ECE		✓		✓	✓		✓	✓	✓	✓
	GSN		✓		✓		✓	✓	✓	✓	✓
	USN		✓			✓	✓	✓	✓	✓	✓
PSNR	–	36.55	37.14	38.68	37.09	38.03	38.33	37.41	37.6	36.77	**39.56**
SSIM	–	0.871	0.884	0.910	0.885	0.901	0.907	0.891	0.894	0.877	**0.955**

and FLOPs by effectively improving the inference speed **2.4 times**. Thus, the experiment results show that our proposed method is effective both in the computational complexity and image quality performance on the practical real noise image dataset. This further proves the effectiveness of the proposed SP and RP methods on the real-world image datasets.

To intuitively demonstrate the superiorities of our method, the results of four widely-used compared methods are visualized in Fig. 6. Obviously, MLP [41] is not apparent in removing complex real noise, while the visual effects of APD-Nets, MIRNet [60] and VDN [44] are better than that of MLP. This is mainly because the powerful convolutional neural networks used in these methods can express the images much more satisfactorily. In all the methods, the visual results produced by our APD-Nets is the closest to that of ground truth. This further demonstrates the effectiveness of the adaptive regularization noise and supplemental prior information for denoising images.

C. Ablation Study

We quantitatively evaluate the effectiveness of all the components in our APD-Nets mainly from two aspects: 1) the SP block; 2) the RP block. All experiments are performed on the SIDD validation dataset.

Ablation study of the SP Block. From Table IV, our SP block increases PSNR and SSIM by **2.42** dB and **0.071** compared to APD-Nets without SP block (Case j vs. Case b). Moreover, each component in the SP block is also vital to improve the performance (Case j vs. Cases g-i). Especially, compared to the regular APD-Nets (Case j), the PSNR and SSIM of the APD-Nets without ERE (Case i) decrease by 2.79 dB and 0.078, respectively. To intuitively illustrate the effectiveness of every component in the SP block, the output SP feature, produced by different combinations of SCA, SSA and ERE, are visualized in Fig. 7. It is clear that, compared to the SP

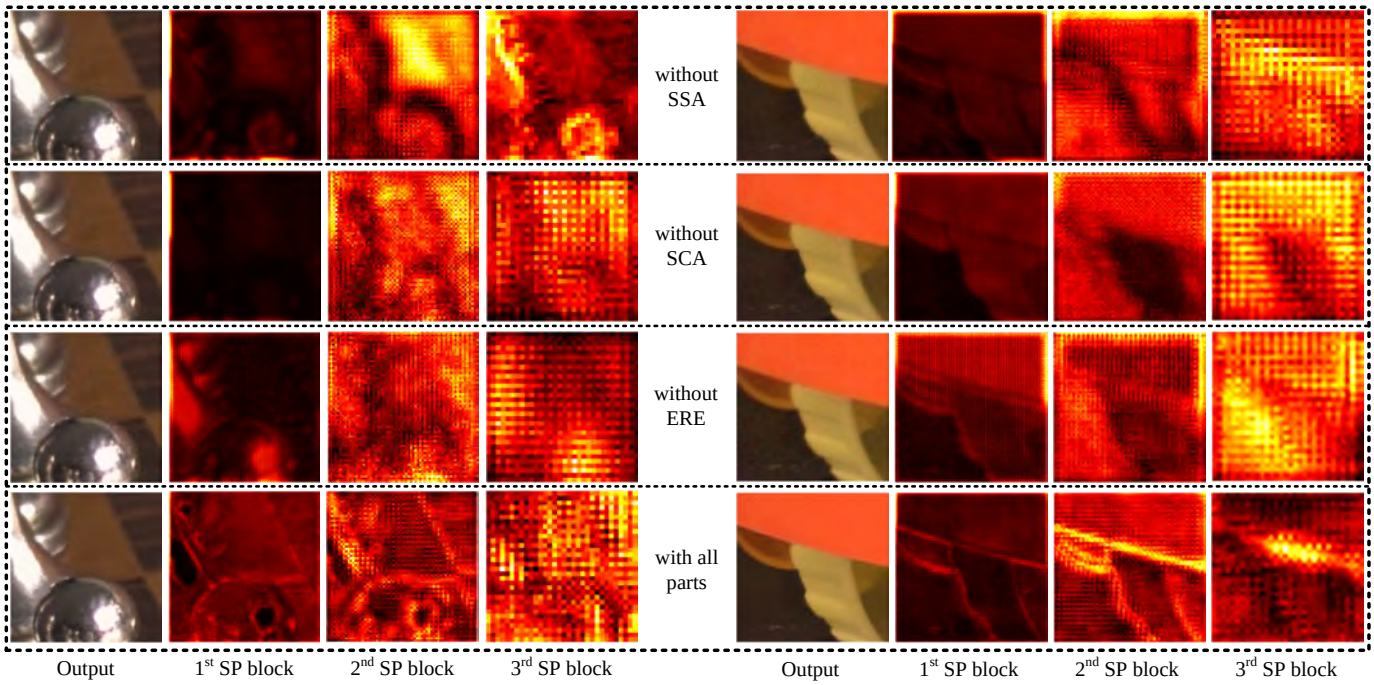


Fig. 7. Visualization effects of SCA, SSA, and ERE in the SP method on SIDD dataset. It is clear that, compared to the SP block with all components, SSA and SCA mainly capture the attention areas, while ERE can learn the corresponding pattern of detail information in every depth to supplement the content in the attention areas for every sample.

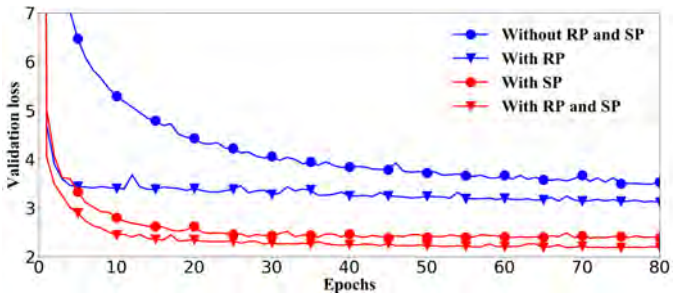


Fig. 8. Comparisons of validation losses on SIDD datasets. Apparently, compared to APD-Nets (without SP and RP) and APD-Nets (with SP), the validation losses are both decreased by APD-Nets (with RP) and APD-Nets (with SP and RP), respectively.

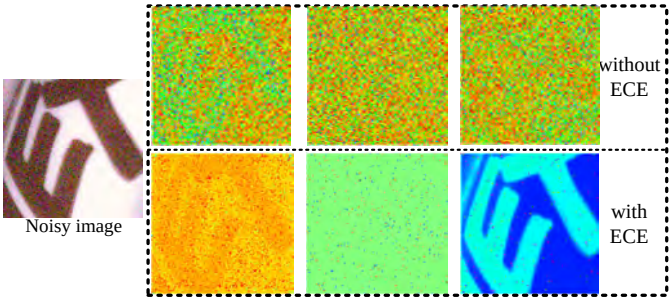


Fig. 9. Visualization effects from ECE, Gaussian noise (G) and Uniform noise (U). ECE can explicitly provide the sample pattern for the generated noise. Gaussian noise can preserve the pattern to some extent, while uniform noise can supplement more scattered noise.

block with all components, SSA and SCA mainly capture the attention areas, while ERE can learn the corresponding pattern of detail information in every depth to supplement the content in the attention areas for every sample. These experiments sufficiently prove every component in SP is essential to produce the adaptive supplemental prior information.

Ablation study of the RP Block. Similarly, from Table IV, our RP block improves the performance compared to the proposed APD-Nets without the RP block (Case j vs. Case c). Furthermore, every component in the RP is effective (Case j vs. Cases d-f), where GSN and USN denote Gaussian noise and Uniform noise, respectively. To prove the regularization effectiveness of RP block, Fig. 8 prints the validation losses from the proposed APD-Nets with various components. (a)Comparison of the verification loss curves of APD-Net without RP and SP block and APD-Net only containing RP block. The detail comparisons are shown in Fig. 8. From this figure, the verification loss curve of the APD-Nets with RP blocks converges faster than that of the APD-Nets without RP and SP blocks, and the former loss value is also lower. This demonstrates the SP block produces a significant effect on the denoising process. (b) Comparison of the verification loss curves of APD-Net including the RP and SP blocks and APD-Net only containing SP blocks. The detail comparisons are shown in Fig. 8. From this figure, the verification loss curve of the APD-Nets including the RP and SP blocks converges faster than that of the APD-Nets only containing SP blocks, meanwhile, the former loss value is also lower. This demonstrates the RP block produces a significant regularization effect on the denoising process. (c) Comparison of all versions of APD-Nets. The detail comparisons are shown

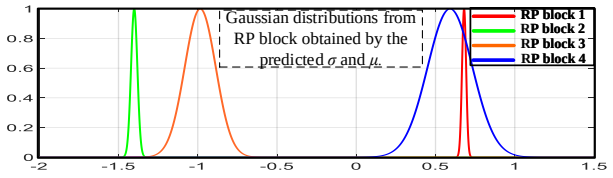


Fig. 10. Learned Gaussian noise distributions from APD-Net's 4 successive RP blocks.

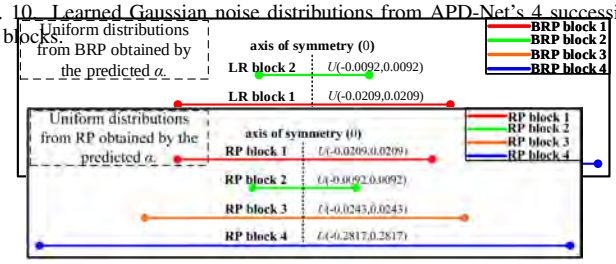


Fig. 11. Learned Uniform noise distributions from APD-Net's 4 successive RP blocks.

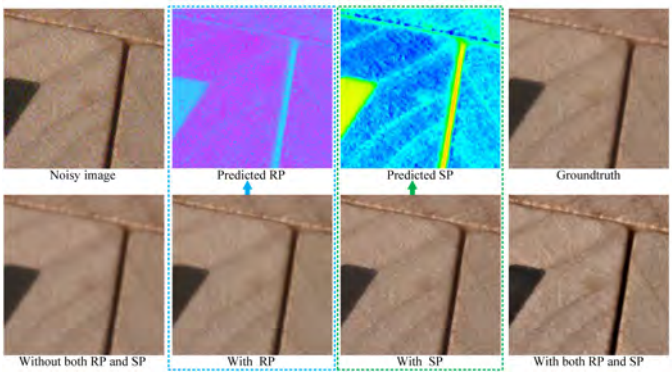


Fig. 12. Visual effect of the proposed APD-Net network containing RP and SP block. The APD-Net that contains both blocks has the best visual effect. The blue arrow indicates the regularized prior information corresponding to the visual results of APD-Nets containing only RP blocks. Furthermore, the green arrow indicates the supplementary prior information corresponding to the visual results of APD-Nets containing only SP blocks.

in Fig. 8. From this figure, the verification loss curve of the APD-Nets including the RP and SP blocks converges fastest and achieves the lowest loss. This shows that the effectiveness of jointly using the RP and SP blocks in our method.

Meanwhile, we visualize the regularization noise produced by our RP block with different components. From Fig. 9, ECE can explicitly provide the sample pattern for the generated noise. Gaussian noise can preserve the pattern to some extent, while uniform noise can supplement more scattered noise. This indicates these components are vital for the adaptive regularization process. Furthermore, we visualized the learned Gaussian noise and Uniform noise distributions in Fig. 10 and 11, respectively. Apparently, the distributions of learned Gaussian noise and Uniform noise produce large differences in all the blocks. This proves the learned noise distributions are significantly adaptive.

Finally, we display the visualized results of image denoising of APD-Net including SP and RP blocks, as shown in Fig. 12. From this figure, it is clear that the generated RP information and SP information are completely different. Specifically, the produced RP information contains obvious noise while the generated SP information contains more detail information, especially for the edges. The relevant recovered image using the SP information also contains more detail information compared to that of using the RP information. Using both RP and SP information can generate most satisfactory images. The comparisons (Case j vs. Case c and Case j vs. Case b) in Table IV also prove using the RP and SP can jointly improve the quality of the recovered images. Noted that, although RP injects the regularization noise into the intermediate feature, it can improve the generalization of the denoised networks. This is beneficial for recovering the noisy images, especially for the real noise images with complex multiple noise resources, improving the quality of the denoised images (the recovered image without RP and SP vs. the recovered image with RP in Fig. 12). Thus, all the above comparisons prove the effectiveness of the proposed RP and SP method.

D. Discussion of Limitations

In this paper, we mainly investigate the internal representation of the proposed APD-Net and the effect of denoising tasks on synthesized and real data. Future research should be undertaken to explore the more tasks, further extended to video denoising. Also, despite the high efficiency of our encoder-decoder design, a limitation remains that we adopt a fixed size learnable matrix, such as ERE and ECE. It may be a better choice to conduct adaptive size based on sizes of low-quality images.

V. Conclusion

This paper proposed APD-Nets from the adaptive prior's view. The proposed APD-Nets is constructed by the Supplemental Priors (SP) block and the Regularization Priors (RP) block. The SP block utilizes ERE, SCA and SSC to generate the supplemental prior information for reconstructing the denoised images and restoring more detail information. The RP block applies ECE, learnable Gaussian noise distribution and Uniform noise distribution to produce the regularization noise for retrieving feature process to prevent the denoising networks from over-fitting. All comparison and visualization experiments have sufficiently proved the effectiveness of the proposed APD-Nets and its every basic component. The study of this work provides a new direction to concurrently improve the generalization of the denoising networks and promote the quality of restored images through learning and inducing adaptive regularization and supplemental priors.

VI. Acknowledgments

This work was supported in part by the NSFC fund 62176077, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019B1515120055, in part by the Shenzhen Key Technical Project under Grant 2020N046, in part by the Shenzhen Fundamental Research Fund under Grant JCYJ20210324132210025, in part by the Guangdong Shenzhen joint Youth Fund under Grant 21201910240005022.

REFERENCES

[1] H. Wang, Y. Li, Y. Cen, and Z. He, "Multi-matrices low-rank decomposition with structural smoothness for image denoising," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 349–361, 2020.

[2] H. Yue, J. Liu, J. Yang, X. Sun, T. Q. Nguyen, and F. Wu, "Ienet: Internal and external patch matching convnet for web image guided denoising," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 3928–3942, 2020.

[3] X. Zhang, J. Zheng, D. Wang, and L. Zhao, "Exemplar-based denoising: A unified low-rank recovery framework," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 2538–2549, 2020.

[4] J. Xu, L. Xu, Z. Gao, P. Lin, and K. Nie, "A denoising method based on pulse interval compensation for high-speed spike-based image sensor," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, pp. 2966–2980, 2021.

[5] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[6] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, pp. 700–711, 2013.

[7] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2862–2869, 2014.

[8] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," *ArXiv*, vol. abs/1807.04364, 2018.

[9] L. Fan, X. mei Li, H. Fan, Y. Feng, and C. ming Zhang, "Adaptive texture-preserving denoising method using gradient histogram and nonlocal self-similarity priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, pp. 3222–3235, 2019.

[10] H. Liu, R. Xiong, D. Liu, S. Ma, F. Wu, and W. Gao, "Image denoising via low rank regularization exploiting intra and inter patch correlation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 3321–3332, 2018.

[11] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Deep image prior," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

[12] L. Enderich, F. Timm, and W. Burgard, "Holistic filter pruning for efficient deep neural networks," *ArXiv*, vol. abs/2009.08169, 2020.

[13] A. Buslaev, A. Parinov, E. Khvedchenya, V. Iglovikov, and A. Kalinin, "Albumentations: fast and flexible image augmentations," *ArXiv*, vol. abs/1809.06839, 2020.

[14] T. Devries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *ArXiv*, vol. abs/1708.04552, 2017.

[15] C. Lee, K. Cho, and W. Kang, "Mixout: Effective regularization to finetune large-scale pretrained language models," *ArXiv*, vol. abs/1909.11299, 2020.

[16] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," *2019 IEEE International Conference on Computer Vision (ICCV)*, pp. 6022–6031, 2019.

[17] C. Gong, T. Ren, M. Ye, and Q. Liu, "Maxup: A simple way to improve generalization of neural network training," *ArXiv*, vol. abs/2002.09024, 2020.

[18] X. Gastaldi, "Shake-shake regularization," *ArXiv*, vol. abs/1705.07485, 2017.

[19] Y. Yamada, M. Iwamura, and K. Kise, "Shakedrop regularization," *ArXiv*, vol. abs/1802.02375, 2018.

[20] K. Zhang, M. Sun, T. Han, X. Yuan, L. Guo, and T. Liu, "Residual networks of residual networks: Multilevel residual networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 1303–1314, 2018.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *ArXiv*, vol. abs/1603.05027, 2016.

[22] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1256–1272, 2017.

[23] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5882–5891, 2017.

[24] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. Huang, "Non-local recurrent network for image restoration," in *2018 Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2018.

[25] C. Chen, Z. Xiong, X. Tian, Z. Zha, and F. Wu, "Real-world image denoising with deep boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 3071–3087, 2020.

[26] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 2305–2318, 2019.

[27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, pp. 3142–3155, 2017.

[28] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, pp. 4608–4622, 2018.

[29] S. Anwar and N. Barnes, "Real image denoising with feature attention," *2019 IEEE International Conference on Computer Vision (ICCV)*, pp. 3155–3164, 2019.

[30] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," *2011 International Conference on Computer Vision (ICCV)*, pp. 479–486, 2011.

[31] J. Xu, L. Zhang, and D. Zhang, "External prior guided internal prior learning for real-world noisy image denoising," *IEEE Transactions on Image Processing*, vol. 27, pp. 2996–3010, 2018.

[32] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," *2009 IEEE International Conference on Computer Vision (ICCV)*, pp. 2272–2279, 2009.

[33] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 60–65 vol. 2, 2005.

[34] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, pp. 2080–2095, 2007.

[35] J. Xu, L. Zhang, D. Zhang, and X. Feng, "Multi-channel weighted nuclear norm minimization for real color image denoising," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1105–1113, 2017.

[36] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *2020 Association for the Advance of Artificial Intelligence (AAAI)*, 2020.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *ArXiv*, vol. abs/1505.04597, 2015.

[38] A. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *NIPS*, 2001.

[39] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 2011–2023, 2020.

[40] I. Higgins, L. Matthey, A. Pal, C. P. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *2017 International Conference on Learning Representations (ICLR)*, 2017.

[41] H. C. Burger, C. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?" *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2392–2399, 2012.

[42] C. Chen, X. Tian, F. Wu, and Z. Xiong, "Udnet: Up-down network for compact and efficient feature representation in image super-resolution," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 1069–1076, 2017.

[43] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4549–4557, 2017.

[44] Z. Yue, H. Yong, Q. Zhao, L. Zhang, and D. Meng, "Variational denoising network: Toward blind noise modeling and removal," in *2019 Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2019.

[45] M. Chang, Q. Li, H. Feng, and Z. Xu, "Spatial-adaptive network for single image denoising," *ArXiv*, vol. abs/2001.10291, 2020.

[46] Y. Kim, J. W. Soh, G. Park, and N. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3479–3489, 2020.

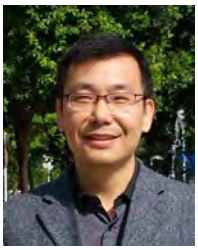
[47] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, pp. 47–57, 2017.

[48] Z. Wang, A. Bovik, H. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.
[49] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140, 2017.
[50] D. Martin, C. C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *2001 IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 416–423 vol.2, 2001.
[51] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, 2017.
[52] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1692–1700, 2018.
[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ArXiv*, vol. abs/1412.6980, 2015.
[54] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.
[55] S. Roth and M. J. Black, "Fields of experts: a framework for learning image priors," *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 860–867 vol. 2, 2005.
[56] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2750–2759, 2017.
[57] S. Nam, Y. Hwang, Y. Matsushita, and S. Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1683–1691, 2016.
[58] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1712–1722, 2019.
[59] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, and M.-H. Yang, "Multi-stage progressive image restoration," *ArXiv*, vol. abs/2102.02808, 2021.
[60] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," *ArXiv*, vol. abs/2003.06792, 2020.

Bo Jiang received the B.Sc. and M.Sc. degree in Electrification and automation, from Northwest Agriculture and Forestry University, China, in 2017 and 2020, respectively. Now he is a Ph.D. student of Harbin Institute of Technology Shenzhen, majoring in Computer Applied Technology. His research interests include deep learning, pattern recognition, computer vision, and low-level image processing.

Yao Lu received the B.S. degree in Software Engineering, Huaqiao University, Xiamen, China, in 2015, and the Ph.D. degree in Computer Applied Technology, Harbin Institute of Technology, Shenzhen, China, in 2020. She was a Post-doctoral Fellow at University of Macau, Macau, China, from 2020 to 2021. She is currently an Assistant Professor with the Biocomputing Research Center, Harbin Institute of Technology, Shenzhen, China. Her research interests include pattern recognition, deep learning, computer vision and relevant applications.

Jiahuan Wang received the B.S. degree from the department of Computer Science and Technology, Huazhong Agriculture University, Wuhan, China, in 2019. She is studying for a M.S. degree at Harbin Institute of Technology, Shenzhen, China. Her research interests include image denoising, deep learning, computer vision and relevant applications.



Guangming Lu received the B.S. degree in electrical engineering, the M.S. degree in control theory and control engineering, and the Ph.D. degree in computer science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2005, respectively. He was a Post-Doctoral Fellow with Tsinghua University, Beijing, China, from 2005 to 2007. He is currently a Professor with the Biocomputing Research Center, Harbin Institute of Technology, Shenzhen, China. He has published over 120 technical papers at prestigious international journals and conferences, including TIP, TNNLS, TCYB, TCSVT, TSMC, CVPR, AAAI, ACMM, IJCAI, etc. His current research interests include pattern recognition, image processing, and automated biometric technologies and applications.

David Zhang received the Ph.D. degree in computer science from Peking University, Beijing, China, the M.Sc. degree in computer science and the Ph.D. degree from the Harbin Institute of Technology (HIT), Harbin, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1994. From 1986 to 1988, he was a Post-Doctoral Fellow with Tsinghua University, Beijing, China, and an Associate Professor with the Academia Sinica, Beijing. He was a Chair Professor from 2005 to 2018 with Hong Kong Polytechnic University, Hong Kong, where he was also the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He is currently a Presidential Chair Professor with School of Data Science, The Chinese University of Hong Kong (Shenzhen), China. He also serves as a Visiting Chair Professor with Tsinghua University and an Adjunct Professor with Peking University, Beijing, Shanghai Jiao Tong University, Shanghai, China, HIT, and the University of Waterloo. His current research interests include medical biometrics and pattern recognition.